

针对科技路线图的文本挖掘研究：集成分析及可视化*

谢秀芳 [1,2]　张晓林 [1]
[1](中国科学院文献情报中心　北京 100190)
[2](首都医科大学卫生管理与教育学院 / 首都医科大学图书馆　北京 100069)

摘要:【目的】实现针对科技路线图内容的知识发现研究，预测未来科技长期发展趋势。【方法】基于采用"抽取–同步–分类"的文本挖掘方法构建的科技路线图信息库，集成分析全球科技发展需求和趋势，对比分析各国发展路线和措施，并以可再生能源领域为案例进行实证研究。【结果】利用开源工具 Timeflow、Gephi 等对实证研究结果进行可视化，按时间序列从多个角度呈现了可再生能源领域到 2050 年的发展态势及各国的战略规划。【局限】综合利用了多种方法工具，自动化程度有待提高，个性化功能有待完善。【结论】该研究方案能够快速获取科技路线图中的核心信息，提高情报获取效率。

关键词: 科技路线图　战略情报　文本挖掘　知识发现　集成分析　信息可视化
分类号: G356.4

1 引　言

科技路线图是关于世界各国未来科技发展规划最直接的战略情报载体，蕴涵着各国科技发展现状、方向、相关技术发展进程、愿景、阶段性目标、战略措施等多方面信息，是一种密集型战略情报资源。利用文本挖掘技术实现对科技路线图的信息抽取、分类组织和集成分析，对把握未来科技发展方向，制定长期发展规划具有重要战略意义[1]。

然而科技路线图在情报学领域更多属于情报研究的产物[2-4]，较少作为情报挖掘的数据资源[5-6]。比较常见的是对某个具体的路线图报告进行人工解读[7-8]，鲜有针对大量科技路线图报告进行文本挖掘研究。因此本研究提出针对科技路线图的文本挖掘研究方案，通过分析科技路线图的内容组织和表达特征，探索科技路线图信息自动抽取方法，建立科技路线图知识库，进而可以基于该特征知识库实现对大量科技路线图的集成分析、对比分析、趋势分析等，实现对科技路线图的文本挖掘和知识发现[9]。

本文基于文献[10]提出的信息抽取方法构建的科技路线图信息库，集成分析全球各国在各个领域的战略规划信息，分析每个领域在世界范围内的发展现状，并预测未来发展趋势，从而实现针对科技路线图文本内容的知识发现过程，为决策者制定相关领域发展计划提供战略情报服务。

2 科技路线图集成对比分析功能框架

本研究前期调研分析了全球 21 个国家或组织发布的 166 份科技路线图，构建了科技路线图内容描述框架和信息分类体系[9]，并在此基础上探索了针对科技路线图文本内容的信息抽取方法[10]，依次进行文本清洗、信息抽取、数据清洗、同步匹配、信息分类等

通讯作者：谢秀芳，ORCID：0000-0002-7405-9844，E-mail：xiexiufang@mail.las.ac.cn。
*本文系中国科学院规划与决策科技支持系统建设项目"科技决策知识服务平台"(项目编号：院1405)的研究成果之一。

步骤，形成了包含文本基础信息、语义分类信息、内容核心信息、句子原始信息 4 类，共计 19 个字段的科技路线图信息库(见图 1)。其中语义分类信息的取值源自信息分类体系[9]中的分类项，如图 2 所示。

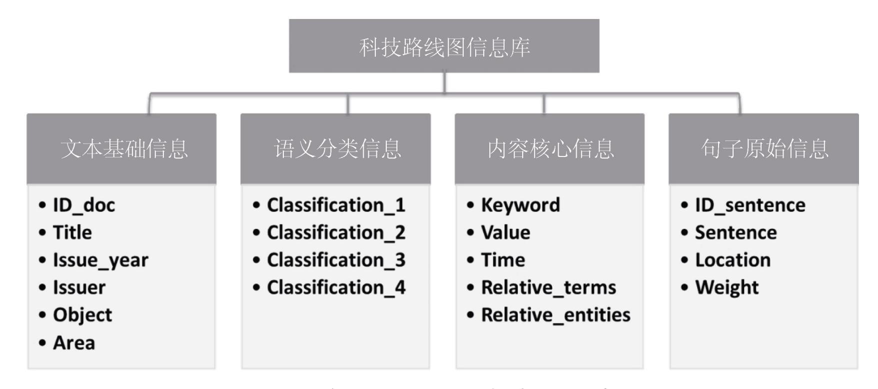


图 1　科技路线图信息库构成字段

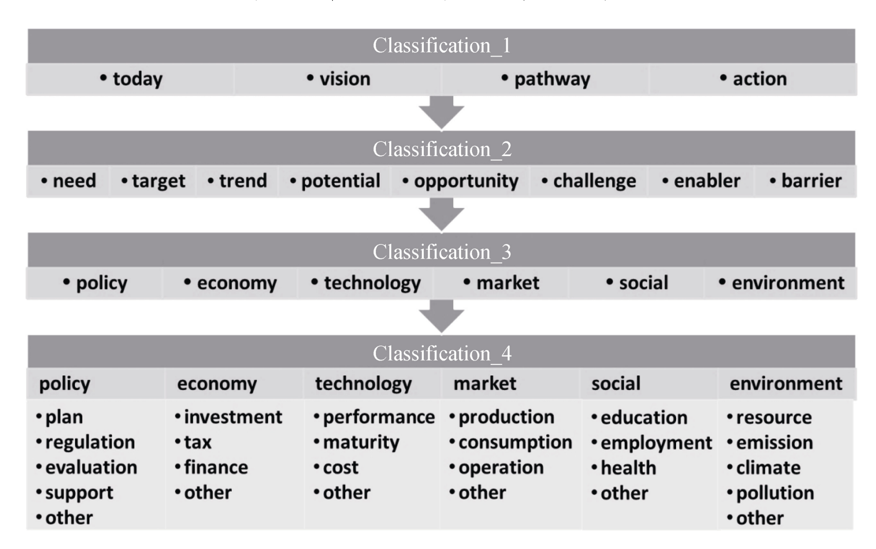


图 2　科技路线图语义分类信息取值

基于该科技路线图信息库，依赖领域(Area)、时间(Time)、关注对象(Object)、语义分类等信息，可以分别从全球和国家两个层面进行如下分析：

(1) 需求分析：集成全球各国对同一领域的发展现状综述，分析该领域在世界范围内的发展需求，以便找准发展定位。

(2) 趋势分析：集成全球各国对同一领域的发展趋势、潜力、机遇的分析论述，预测该领域在世界范围内的长期发展趋势，及时把握发展机遇。

(3) 路线分析：集成对比分析全球各国在同一领域的技术发展路线，为制定符合国情的发展路线提供参考服务，如：

①技术发展态势分析：集成全球各国在同一领域不同时段的技术选择信息，分析该领域的技术发展态势；

②技术发展方向分析：集成全球各国在同一领域不同时段的技术目标信息，预测该领域的未来技术发展方向；

③技术发展路径分析：集成同一技术在全球各国的发展目标，对比各国针对相同技术的不同发展路径规划。

(4) 战略分析：集成对比分析全球各国在同一领域采取的战略措施，为决策者适时适当的战略布局提供战略情报服务，如：

①不同国家的发展愿景分析：在相同时间区间，对比分析不同国家在同一领域的不同发展愿景目标；

②不同国家的发展路线分析：在相同时间区间，对比分析不同国家在同一领域的不同发展路线规划；

③不同国家的发展措施分析：在相同时间区间，对比分析不同国家在同一领域采取的不同发展战略措施。

因为科技路线图通常是由权威机构或部门针对相关领域发布的未来发展计划，所以具有一定的权威性

和前瞻性，通过上述分析，可以比较全面地了解相关领域在世界范围内的发展现状和趋势，该领域未来发展的关键技术，以及各国在该领域的技术选择和战略布局，因此对把握发展机遇、制定发展战略具有重要的情报价值和参考意义。

3 科技路线图集成分析实现方法

3.1 需求分析

为了分析某个领域在世界范围内的发展需求，集成全球各国对该领域发展现状的调查结果，本研究使用开源工具 Timeflow[11]进行可视化呈现，将属于同一领域的现状(Classification_1="today")及需求(Classification_2 ="need")信息在时间(Time)轴上分类(Classification_3)呈现，并用权重(W)标识关键词节点大小，用不同颜色标识分类指标(Classification_4)。为了实现上述功能，需对已有数据进行如下预处理:

(1) 关键词与时间词的关联匹配

由于同一个句子中可能包含多个关键词和时间词，所以需要将每个关键词和时间词关联匹配，本研究使用 Python 编码，遍历同一个句子内的每个关键词和时间词，借助值词和二者的个数关系实现一对一匹配，详情可以参考文献[10]的核心信息同步匹配部分。

(2) 时间词的赋值计算

为了将各类信息在时间轴上呈现，针对以"today/year/decade/century 等"方式表达的时间词，统一映射为数字表达方式，同时为"Time"字段为空的数据记录根据 Classification_1 字段结果赋予时间信息。设报告发布年份等于 t，以 t 为分界点，对包含"year/decade/century 等"的时间词按表 1 中规则分别取计算符号和数值，计算出对应的时间信息"t +/- n"。对其他不涉及计算的时间词按表 2 中规则直接进行赋值。

表 1 时间词计算规则

规则	描述	操作
计算符号(+/-)的判定规则(决定在时间 t 的基础上执行"加"或"减"操作)	过去时间	over/past/previous/recent/preceding/last 等，符号取"-"
	将来时间	next/following/coming/later 等，符号取"+"
计算加数或减数的取值规则(在时间 t 的基础上执行加减的数值 n)	单数	year: n=1; decade: n=10; century: n=100
	two/three/ten/20 等后接复数	years: 取相应数字，即 n=2/3/10/20 等 decades: 取相应数字乘以 10，即 n=10×(2/3/10/20)等 centuries: 取相应数字乘以 100，即 n=100×(2/3/10/20)等
	few/several/some 等后接复数	years: n=5; decades: n=50; centuries: n=500

表 2 时间词赋值规则

条件		赋值
当前时间(如 today/present/current/now)		赋值为发表年份 t
beginning/early/dawn/start of century/decade		赋值为该 10 年或世纪开始的年份
end/late of decade/century		赋值为该 10 年或世纪结束的年份
时间词为空(不包含且未能继承时间信息)	Classification_1="today"	即属于现状，赋予发表年份 t
	Classification_1="vision"	即属于愿景，赋予最后一个规划节点年份
	Classification_1="pathway"/"action"	即属于路线或措施，赋予整个规划区间
	Classification_1="other"	即属于其它，赋予 t–1，作为其它背景信息

虽然 Timeflow 可以呈现时间区间，为了使呈现结果更加清晰可辨，本研究中将时间区间分解为多个时间规划节点，如区间(2015, 2050)分解为"2015、2020、2030、2050", (2011, 2030)分解为"2011、2020、2030"。然后使所在句子的关键词一一与每个时间节点关联匹配。

(3) 关键词权重计算

根据关键词在每篇文档中所处位置(Location)、词频(TF，在某篇文档内出现的次数)和文档频次(DF，在语料库中出现该关键词的文档数)赋予不同的权重。

①位置权重(w_p): 根据位置重要性从高到低赋予其不同的权重值[9, 12], w_{head}=7、w_{title}=6、w_{item}=5、w_{begin}=4、w_{lead}=3、w_{end}=2、w_{plain}=1;

②词频权重(w_t): 取关键词词频的自然对数 w_t=ln *TF*;

③文档频次权重(w_d): 取关键词文档频次的自然对数 w_d= ln *DF*。

关键词总权重 $W = w_p + w_t + w_d$, 其中 w_p 和 w_t 反映关键词在所属文档中的局部重要性, w_d 反映该关键词在语料库中的全局重要性, 综合权重 *W* 反映了该关键词在整个领域发展规划中的重要性。

3.2　趋势分析

集成全球各国对未来某个领域发展趋势、潜力、机遇的判断分析, 可以预测该领域在世界范围内的总体长期发展趋势, 本研究使用开源工具 Timeflow 进行可视化呈现, 将属于同一领域的愿景(Classification_1 =“vision”)及趋势(Classification_2 =“trend & potential & opportunity”)信息在时间(Time)轴上分类(Classification_3)呈现, 并用权重标识关键词节点大小, 用不同颜色标识分类指标(Classification_4)。其中时间和权重计算方法同需求分析。

3.3　路线分析

为了预判某个领域的技术发展走向, 集成全球各国在该领域的未来发展路线规划信息, 本研究使用 Gephi[13], 将各国在同一领域的技术选择及其目标按时间规划节点呈现, 关键词作为节点标签, 关键词之间的句内共现关系作为边。因为内容核心信息中的关联术语词是所在句子包含的全部术语关键词, 因此可以利用同一句子编码的关键词和关联术语词两个字段构建边关系, 本研究采用 Python 编程实现关键词网络的构建, 主要算法步骤如下。

(1) 技术发展态势分析

①筛选某个领域(Area), 集成分析各国路线(Classification_1=“pathway”)规划中各时间阶段(Time)的技术(Classification_3=“technology”)发展趋势、潜力、机遇(Classification_2=“trend & potential & opportunity”)的关键词(kw_i)、权重(w_i)和关联术语(rt_j)信息;

②构建关键词网络, 遍历每个关键词 kw_i:

　　if 已存在节点 $Node_{kw_i}$

　　　　$Node_{kw_i}$ 的权重 $w_{kw_i} = w_{kw_i} + w_i$

　　else

　　　　新增节点 $Node_{kw_i}$, 权重 $w_{kw_i} = w_i$

　　　　遍历每个关联术语词 rt_j

　　if 已存在 kw_i 到 rt_j 的边 $Edge_{kw_i, rt_j}$

　　　　$Edge_{kw_i,rt_j}$ 的权重 $w_{kw_i,rt_j} = w_{kw_i,rt_j} + 1$

　　else

　　　　新增 $Edge_{kw_i,rt_j}$ 的权重 $w_{kw_i,rt_j} = 1$

③输出每个节点 $Node_{kw_i}$ 及其对应的权重 w_{kw_i} 到文件 node_trend.txt, 输出每条边 $Edge_{kw_i,rt_j}$ 及其权重 w_{kw_i,rt_j} 到文件 edge_trend.txt;

④将 node_trend.txt 和 edge_trend.txt 导入 Gephi, 用节点权重 w_{kw} 标识节点标签大小和颜色深浅, 边权重 $w_{kw,\ rt}$ 标识边的粗细和颜色深浅。

(2) 技术发展方向分析

①筛选某个领域(Area), 集成分析各国路线(Classification_1=“pathway”)规划中的技术(Classification_3=“technology”)发展目标(Classification_2=“need & target”)的关键词、权重(Weight)和关联术语(Relative_terms)信息;

②构建关键词网络文件并导入 Gephi 可视化, 方法与态势分析相同。

(3) 技术发展路径分析

筛选某个技术关键词(Keyword), 对比分析全球各国针对该项技术发展路径(Classification_1=“pathway”)的不同目标(Classification_2=“target”)规划, 使用 Timeflow 按时间轴(“Time”)呈现不同国家(“Object”)的发展路径。

3.4　战略分析

为对比分析全球各国在同一领域的战略规划, 如发展愿景、路线和采取的相应措施, 使用开源工具 Timeflow 对所关注国家在特定领域的相关战略信息进行全面可视化分析, 并按时间序列(以“time”为横轴)分类(以“Classification_3”为纵轴)呈现, 用权重(W)标识关键词节点大小, 用不同颜色标识分类指标(Classification_4)。选择某个领域(Area)和关注对象(Object), 集成其在政策、经济、技术、市场等各方面(Classification_3)的信息, 分析关注对象在该领域的发展愿景、路线及措施, 具体分析目标及其实现配置如表 3 所示。

表 3　针对关注对象在某个领域的战略分析参数设置

功能目标	参数设置	公共配置
发展愿景分析	Classification_1=“vision”; Classification_2=“target”;	x 轴: time; y 轴: Classification_3;
发展路线分析	Classification_1=“pathway”; Classification_2=“target”;	权重: Weight; 颜色: Classification_4;
发展战略分析	Classification_1=“action”; Classification_2=“target & need & barrier & enabler”;	节点: Keyword; 详情: 鼠标悬停节点显示原句及所属全部信息字段;

4　科技路线图集成分析结果可视化

本研究以可再生能源领域(Renewable Energy)为例，集成全球针对该领域发布的科技路线图信息，分析可再生能源领域的发展现状和趋势，未来发展的关键技术，以及不同国家在该领域的技术选择和拟采取的战略措施。

4.1　需求分析示例

集成各国在可再生能源领域的现状需求信息，并利用 Timeflow 进行可视化，如图 3 所示，可以看出该领域的发展需求包含技术(Technology)、市场(Market)、政策(Policy)、环境(Environment)和经济(Economy)方面，且主要是技术、市场和政策需求。

(1) 技术方面：主要是性能(performance，橙色)和成熟度(maturity，品红色)，如：提高太阳能集热器的性能、先进生物燃料的商业化等;

(2) 市场方面：主要是生产(production，红色)和消费(consumption，深红色)，如：增加发电量、可再生能源份额、能源需求量等;

(3) 政策方面：主要是规划(plan，绿色)和支撑(support，蓝色)，如加强光伏研究、制定国家能源计划、增加水电站项目等。

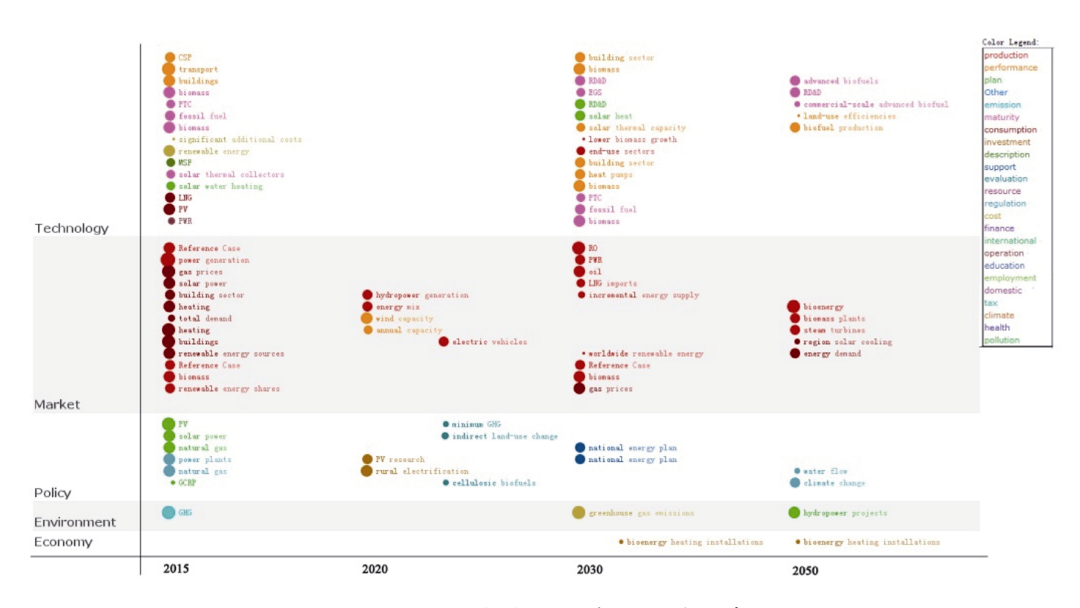


图 3　可再生能源领域的发展需求

(注：参数设置 Classification_1 =“today”, Classification_2 =“need”, Year≥2015)

4.2　趋势分析示例

集成各国在可再生能源领域的发展愿景趋势信息，利用 Timeflow 可视化如图 4 所示，可以看出该领域的发展趋势包含市场(Market)、技术(Technology)、环境(Environment)、政策(Policy)、经济(Economy)、社会(Social)等方面，且主要是技术、市场和环境趋势。

(1) 市场方面：主要是生产(production，红色)和消费(consumption，深红色)，如：增加可再生能源份额、发电量，降低价格、增量成本等;

(2) 技术方面：主要是性能(performance，橙色)和成本(cost，黄褐色)，如：内置热存储、光伏发电、发电厂效率、降低投资成本等;

(3) 环境方面：主要是排放(emission，青色)和资源(resource，紫色)，如减少温室气体排放、扩大可再生资源的开发等。

4.3　路线分析示例

(1) 技术发展态势

集成全球可再生能源领域的技术(Technology)发展趋势(Trend)、潜力(Potential)、机遇(Opportunity)等信息，利用 Gephi 进行可视化，如图 5 所示，呈现了 2015 至 2050 年可再生能源领域的技术发展态势。从图中可以看出该领域的发展态势主要围绕聚焦式太阳能(CSP)、生物质能(biomass)、光伏(PV)、风能(wind power)等技术，包括技术成本(cost)、碳排放(CO_2)、发电量(power)等方面的性能提升。

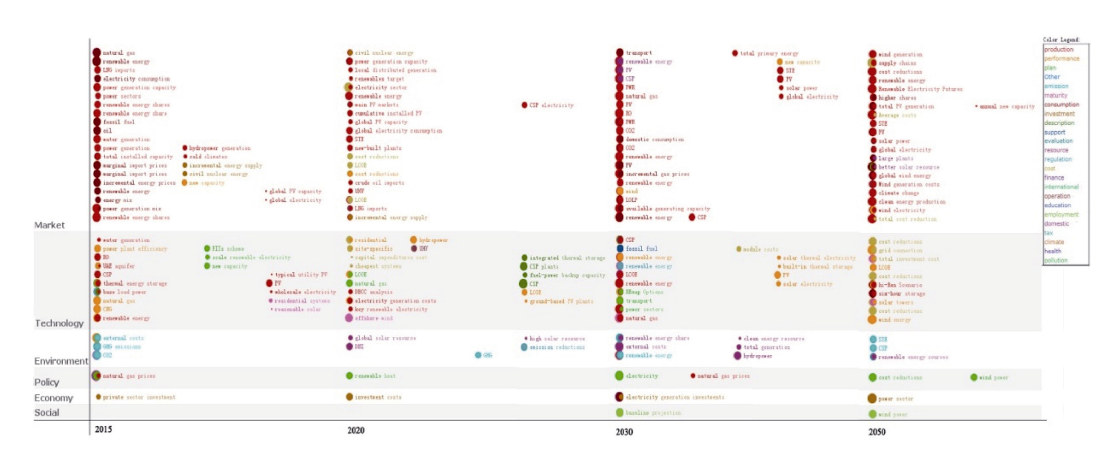


图 4　可再生能源领域的发展趋势

(注：参数设置 Classification_1 =“**vision**”, Classification_2 =“trend & potential & opportunity”, Year≥2015)

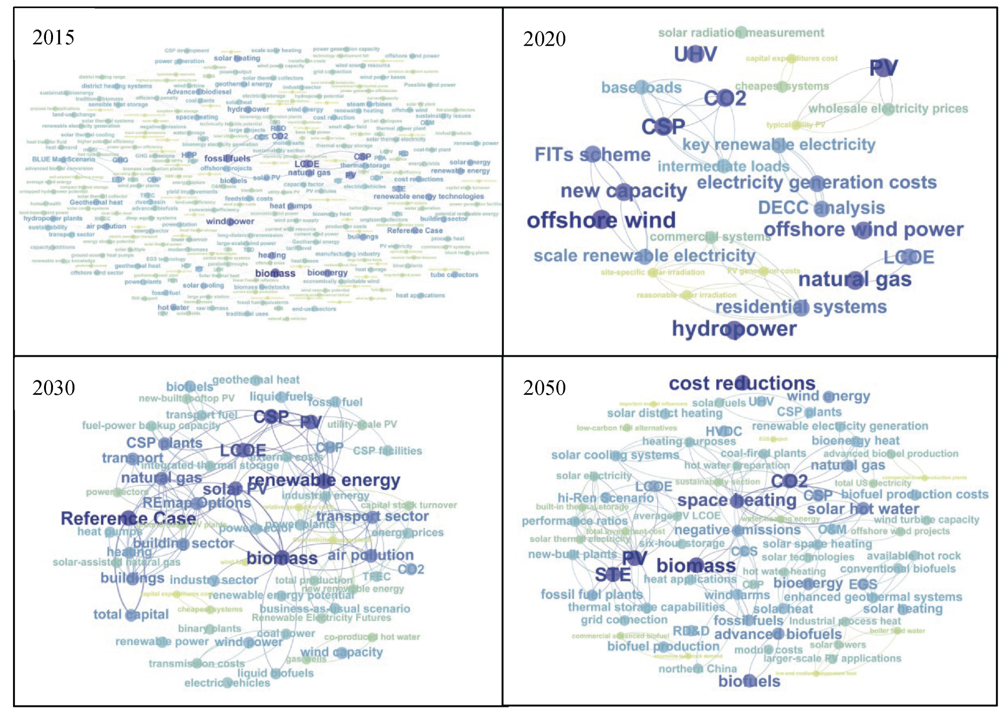


图 5　可再生能源领域的技术发展态势

(2) 技术发展方向

集成全球各国可再生能源领域的技术(Technology)发展目标(Target)信息，利用 Gephi 进行可视化，如图 6 所示，呈现了 2015 至 2050 年可再生能源领域的技术发展方向。从图中可以看出该领域的技术发展方向主要集中在聚焦式太阳能发电(CSP)、光伏(PV)、水电(hydropower)、风电(wind power)、生物质能(biomass)、生物燃料(biofuel)等技术在交通(transport)、建筑(building)、热力(heating)等方面的应用。

(3) 技术发展路径

选则某个技术关键词，对比分析全球各国针对该项技术发展路径(pathway)的不同目标(target)规划，用 Timeflow 进行可视化。图 7 呈现了可再生能源领域 4 种主要技术：聚焦式太阳能发电技术(CSP)、碳捕获与储存技术(CCS)、风电技术(wind power)和光伏技术(PV)，在不同国家的发展路径规划。以 CSP 为例，鼠标悬停在每个节点，可以了解该技术在该时间点的相关发展状态信息，对比各国该项技术的发展路径。

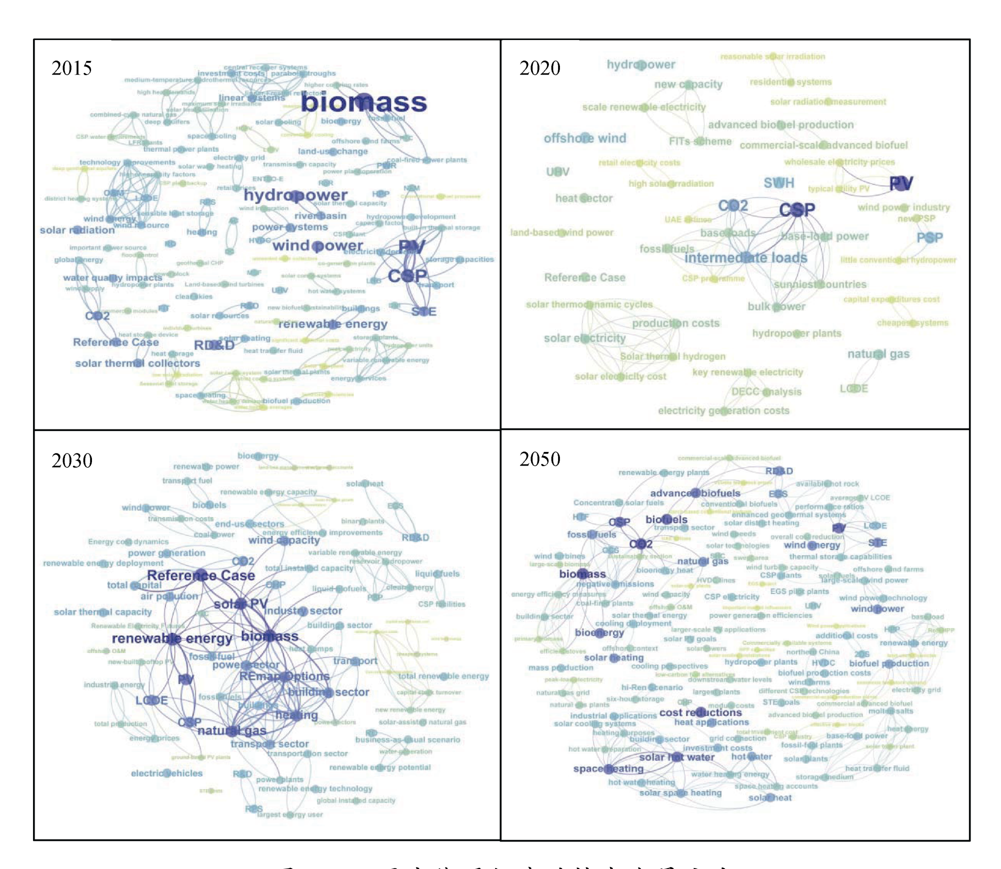


图 6　可再生能源领域的技术发展方向

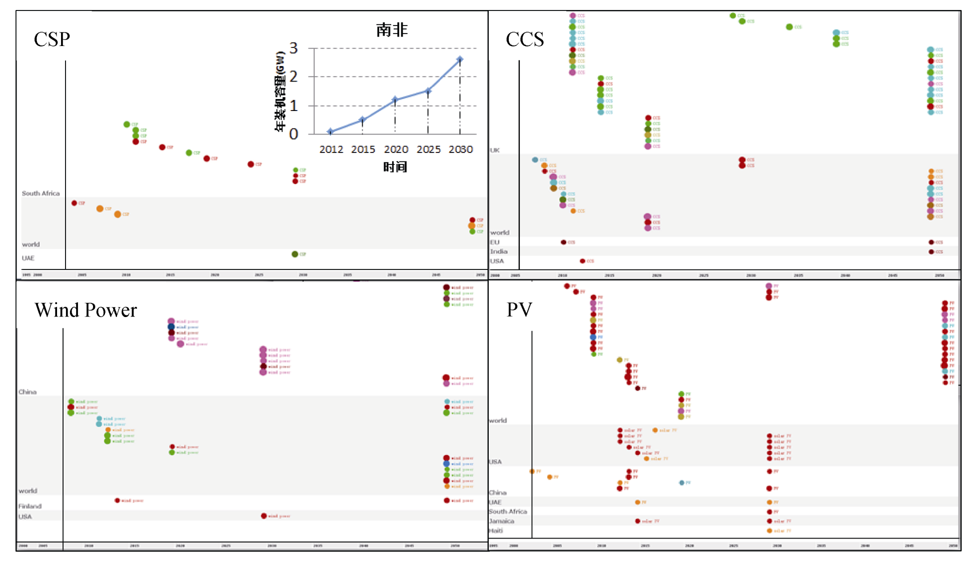


图 7　可再生能源领域的技术发展路径

①南非(South Africa): 太阳能装机容量 2012 年 100MW→2015 年 500MW→2020 年 1.2GW→2025 年 1.5GW →2030 年 2.6GW;

②全球(World): 2050 年太阳能装机容量 650GW, 发电量达到 2200TWh, CO2 排放减少 50%;

③阿联酋(UAE): 2030 年降低 CSP 成本;

4.4　战略分析示例

对比分析美国和中国在可再生能源领域的发展愿景、路线和措施, 使用 Timeflow 对两国在该领域的相关战略信息进行全面可视化分析, 如图 8 所示。

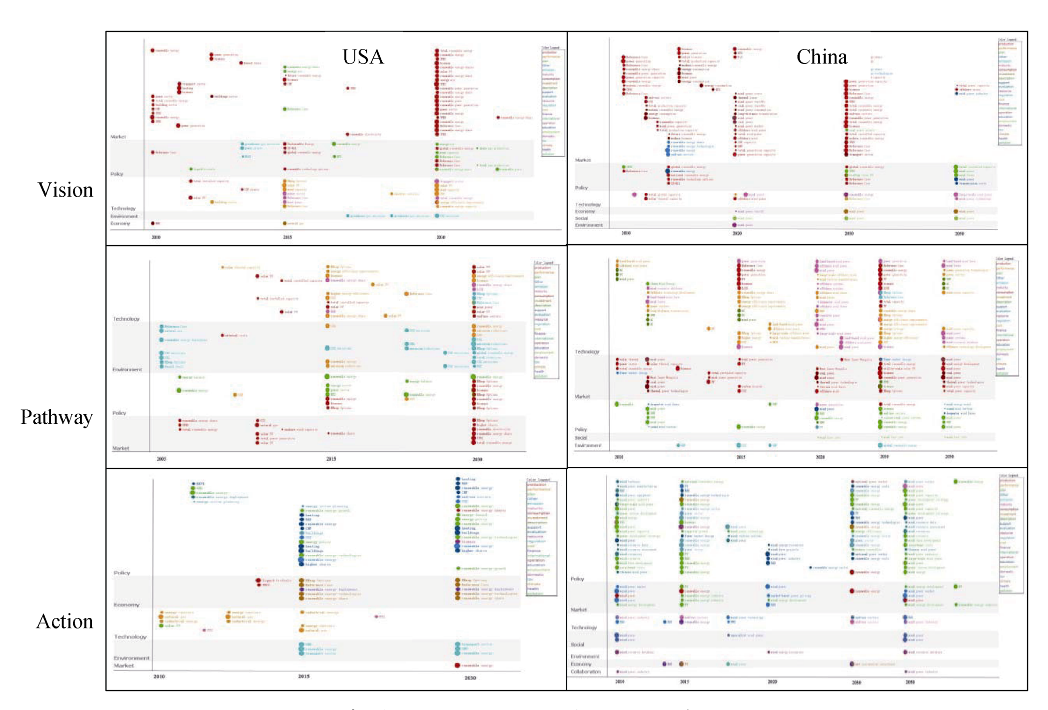


图 8 中美可再生能源领域的发展战略对比

(1) 发展愿景对比

①美国在可再生能源领域的发展愿景，包含市场、政策、技术、环境、经济等方面，且主要是市场、政策和技术目标。

1)市场方面：主要是生产(production，红色)，如：增加可再生能源发电量、扩大可再生能源份额等；

2)政策方面：主要是规划(plan，绿色)，如：可持续发展能源计划、修建可再生能源发电厂等；

3)技术方面：主要是性能(performance，橙色)，如提高太阳能、风能效率等。

②中国在可再生能源领域的发展愿景，包含市场、政策、技术、经济、社会、环境等方面，且主要也是市场、政策和技术目标。

1)市场方面：主要是生产(production，红色)和消费(consumption，深红色)，如：增加可再生能源发电量、装机容量，风电价格低于煤电价格，降低风力发电成本等；

2)政策方面：主要是规划(plan，绿色)和支持(support，蓝色)，如：制定风力发展战略、激励电网公司购买再生电力等；

3)技术方面：主要是成熟度(maturity，品红色)，如提高风力发电技术、大型风电场、海上风电场技术等。

(2) 发展路线对比

①美国在可再生能源领域的发展路线，包含技术、环境、政策和市场。

1)技术方面：主要是性能(performance，橙色)，如提高风能、太阳能效率；

2)环境方面：主要是排放(emission，青色)，如相比 2005 年，减少 CO_2 排放→2020 年(17%)→2025 年(26%~28%)→2030 年(33%)→2050 年(80%)；

3)政策方面：主要是规划(plan，绿色)，如制定加速可再生能源发展的路线；

4)市场方面：主要是生产(production，红色)，如：增加可再生能源份额 2010 年~2030 年→26%，增加向岸风能装机容量 2014 年(63GW)→2030 年(314GW)。

②中国在可再生能源领域的发展路线，包含技术、市场、政策、社会、环境等方面，且主要是技术、市场和政策。

1)技术方面：主要是性能(performance，橙色)和成熟度(maturity，品红色)，如 2010 年~2015 年风电容量 15GW→2020 年风电容量 200GW→2030 年风电容量 400GW→2050 年风电容量 1TW，2010 年~2020 年陆地风电→2020 年~2030 年近海风电→2030 年~2050 年远海风电和微观选址技术；

2)市场方面：主要是生产(production，红色)和消费(consumption，深红色)，如：2020 年风电替代 130mtce(百万吨煤当量)→2030 年风电可替代 260mtce→2050 年风电可替代 660mtce；

3)政策方面：主要是规划(plan，绿色)，如制定可再生能源发展战略。

(3) 发展措施对比

①美国在可再生能源领域的发展措施，包含政策、经济、技术、环境、市场等方面，且主要是政策、经济和技术措施。

1)政策方面：主要是支撑(support，蓝色)，如：支持可再生能源技术的研发；

2)经济方面：主要是投资(investment，棕色)，如：每年为可再生能源技术投入860亿美元；

3)技术方面：主要是性能(performance，橙色)，如：提高能源载体、太阳能光伏性能。

②中国在可再生能源领域的发展措施，包含政策、市场、技术、社会、环境、经济、合作等方面，且主要是政策、市场措施。

1)政策方面：包括规划(plan，绿色)、管理(regulation，蓝绿色)、评价(evaluation，深蓝色)、支撑(support，蓝色)，如：制定可再生能源工业发展规划、电价管理法规等，提高风电技术标准、测试和认证体系，提供技术咨询、战略研究等服务；

2)市场方面：主要是市场机制，加快电力市场改革、建立风电管理和市场规则、电力定价市场化等。

5 结 语

本研究采用前期提出的文本挖掘研究方案及信息抽取方法，基于已构建科技路线图信息库，以可再生能源领域为案例集成分析世界范围内该领域的发展需求和长期发展趋势，对比分析各国技术发展路线、战略发展措施等，并利用Timeflow、Gephi等开源工具进行可视化呈现。实验结果表明，本研究方法切实可行，能够较快速地把握全球发展现状和趋势，较全面地掌握世界各国未来的长期发展路线和措施，为决策者制定符合国情的发展规划提供较全面、快速的战略情报服务。

研究尚处于方法探索阶段，仍存在不足有待改进：综合利用了多种编程语言和软件工具，受限于笔者的时间和精力，暂未实现所有工具的整合，因而自动化程度有待进一步提高；可视化均选择了开源工具的已有功能，在未来的研究中可以根据具体需求灵活订制功能，例如目前仅能实现鼠标悬停在关键词节点显示相关信息，可以增加点击关键词进入关联段落、篇章等详情页面，更多功能有待完善。

参考文献：

[1] 刘细文，柯春晓.技术路线图的应用研究及其对战略情报研究的启示[J]. 图书情报工作，2007, 51(6): 37-40, 112. (Liu Xiwen, Ke Chunxiao. The Applications of Technology Roadmap and Its Enlightenment to Strategic Intelligence Research [J]. Library and Information Service, 2007, 51(6): 37-40, 112.)

[2] Zhang Y, Zhang G, Chen H, et al. Topic Analysis and Forecasting for Science, Technology and Innovation: Methodology with a Case Study Focusing on Big Data Research [J]. Technological Forecasting and Social Change, 2016, 105: 179-191.

[3] Amer M, Daim T U, Jetter A. Technology Roadmap Through Fuzzy Cognitive Map-Based Scenarios: The Case of Wind Energy Sector of a Developing Country [J]. Technology Analysis & Strategic Management, 2016, 28(2): 131-155.

[4] Jin G, Jeong Y, Yoon B. Technology-driven Roadmaps for Identifying New Product/Market Opportunities: Use of Text Mining and Quality Function Deployment [J]. Advanced Engineering Informatics, 2015, 29(1): 126-138.

[5] 叶春蕾，冷伏海. 技术路线图中未来技术词表构建方法研究[J]. 现代图书情报技术，2013(5): 59-63. (Ye Chunlei, Leng Fuhai. Building the Future—Oriented Technology Thesaurus of Technology Roadmap[J]. New Technology of Library and Information Service, 2013(5): 59-63.)

[6] 叶春蕾，冷伏海. 基于词汇链的路线图关键词抽取方法研究[J]. 现代图书情报技术，2013(1): 50-56. (Ye Chunlei, Leng Fuhai. Study on the Keyword Extraction from Roadmap Based on the Lexical Chains [J]. New Technology of Library and Information Service, 2013(1): 50-56.)

[7] Amer M, Daim T U. Application of Technology Roadmaps for Renewable Energy Sector [J]. Technological Forecasting and Social Change, 2010, 77(8): 1355-1370.

[8] Bader B, Richardson C, Tsuriya M. Technology Roadmap Overviews and Future Direction through Technology Gaps[C]// Proceedings of the 2015 International Conference on Electronics Packaging. 2015.

[9] 谢秀芳，张晓林. 针对科技路线图的文本挖掘研究框架及特征分析[J]. 情报科学. (待发). (Xie Xiufang, Zhang Xiaolin. Text-mining Framework and Feature Analysis on Science and Technology Roadmap [J]. Information Science. In Press.)

[10] 谢秀芳，张晓林. 针对科技路线图的文本挖掘研究：信息抽取方法[J]. 情报理论与实践. (待发). (Xie Xiufang, Zhang Xiaolin. The Research on Text-mining of Science and Technology Roadmap: Method of Information Extraction[J]. Information Studies: Theory & Application. In Press.)

[11] Timeflow [EB/OL]. [2016-07-22]. https://github.com/FlowingMedia/TimeFlow/wiki.

[12] 史磊，王永成. 英文文献自动摘要系统的研制与开发[J]. 高技术通讯，1999, 9(11): 22-26. (Shi Lei, Wang Yongcheng. Research and Development of an Automatic Abstracting System for English Documents [J]. Chinese High Technology Letters, 1999, 9(11): 22-26.)

[13] Gephi [EB/OL]. [2016-07-22]. https://gephi.org/.

作者贡献声明:

谢秀芳: 设计研究方案, 采集、清洗和分析数据, 撰写论文;
张晓林: 确定研究方向, 提出研究思路, 修改论文。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: xiexiufang@mail.las.ac.cn。

[1] 谢秀芳, 张晓林. Keywords-network.py. 关键词网络构建代码.
[2] 谢秀芳, 张晓林. data-20160608-timesplit.csv. 需求分析(图 3)、趋势分析(图 4)、路径分析(图 7)、战略分析(图 8)支撑数据集.
[3] 谢秀芳, 张晓林. Keywords_trend.rar、node_trend.rar、edge_trend.rar. 技术发展态势分析支撑数据(图 5).
[4] 谢秀芳, 张晓林. Keywords_VsTg.rar、node_VsTg.rar、edge_VsTg.rar. 技术发展方向分析支撑数据(图 6).

收稿日期: 2016-09-30
收修改稿日期: 2016-11-02

Integrated Analysis and Visualization of Sci-Tech Roadmaps: Case Study of Renewable Energy

Xie Xiufang[1, 2]　Zhang Xiaolin[1]
[1](National Science Library, Chinese Academy of Sciences, Beijing 100190, China)
[2](School of Health Management and Education, Capital Medical University/
Library of Capital Medical University，Beijing 100069, China)

Abstract: [**Objective**] This study aims to predict the development trends of science and technology (S&T) with knowledge extracted from S&T roadmaps (STR). [**Methods**] First, we constructed an STR information database based on the “extraction – synchronization – classification” method of text mining. Second, we analysed the demands and trends of global S&T progress. Finally, we compared and analyzed different countries’ S&T strategies in the field of renewable energy. [**Results**] We used open source tools, such as Timeflow, Gephi to visualize the results of this case study, such as the globle development trends and national strategic planning in the field of renewable energy by 2050. [**Limitations**] The automation and personalization features of this study need to be improved. [**Conclusions**] The proposed method could retrieve strategic intelligence from the STRs effectively.

Keywords: Science and Technology Roadmap　Strategic Intelligence　Text Mining　Knowledge Discovery　Integrated Analysis　Information Visualization